

Korrekturverfahren zur Berechnung der Einkommen über der Beitragsbemessungsgrenze

Jan Binder, Barbara Schwengler

Korrekturverfahren zur Berechnung der Einkommen über der Beitragsbemessungsgrenze

Jan Binder (STASA), Barbara Schwengler (IAB)

Auch mit seiner neuen Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

Also with its new series "IAB Discussion Paper" the research institute of the German Federal Employment Agency wants to intensify dialogue with external science. By the rapid spreading of research results via Internet still before printing criticism shall be stimulated and quality shall be ensured.

Inhaltsverzeichnis

Abstract	4
1 Einleitung	5
1.1 Der Bruttojahreslohn pro Beschäftigtem im Jahreszeitraum- material	5
1.2 Beitragsbemessungsgrenze der sozialversicherungspflichtigen Einkommen	7
2 Korrekturverfahren	10
2.1 Allgemeines Vorgehen	10
2.2 Eigenschaften der Log-Normalverteilung	11
2.3 Anpassung der Log-Normalverteilung an die Einkommens- verteilung	12
3 Weitere Ergebnisse des Korrekturverfahrens	17
4 Fazit	21
Literatur	22
Anhang	24

Abstract

Der durchschnittliche Bruttojahreslohn pro Beschäftigtem in einer Region liefert wertvolle Informationen über die regionalen Verdienstmöglichkeiten. Bedingt durch das Meldeverfahren an die Sozialversicherungsträger werden die Einkommen jedoch nicht mit ihrer vollen Höhe erfasst, sondern nur bis zur jeweils gültigen Beitragsbemessungsgrenze. Alle Einkommen über der Beitragsbemessungsgrenze werden abgeschnitten. Der in einer Region gemessene durchschnittliche Bruttojahreslohn pro Beschäftigtem aller Beschäftigungsverhältnisse liegt somit niedriger als der „tatsächliche“ Mittelwert. Um die realen Verdienstmöglichkeiten in den einzelnen Regionen dennoch möglichst exakt abzubilden, muss dieser durch das institutionelle Abschneideverfahren bedingte Fehler bereinigt werden. Dazu wird die Einkommensverteilung über der Beitragsbemessungsgrenze sowie die „abgeschnittene“ Bruttolohn- und Gehaltssumme hinzugeschätzt und damit der Mittelwert für jede Region einzeln korrigiert. Das Verfahren dieser Lohnkorrektur wird in dem vorliegenden Beitrag ausführlich beschrieben.

1 Einleitung

Der durchschnittliche Bruttojahreslohn pro Beschäftigtem in einer Region ist eine Messgröße, die Hinweise auf die regionalen Verdienstmöglichkeiten liefert. Aus diesem Grund wurde dieser Indikator sowohl in der Vergangenheit als auch in der Gegenwart neben anderen regionalen Größen, wie z. B. der Arbeitslosenquote, der Erwerbstätigenprognose und dem Infrastrukturindikator, zur Messung der Wirtschaftskraft von Regionen für die Abgrenzung der Fördergebiete der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ (GA) herangezogen (Koller, Schwengler 2000, Deutscher Bundestag 2000, Schwengler et al. 2006).

1.1 Der Bruttojahreslohn pro Beschäftigtem im Jahreszeitraummaterial

Datengrundlage für den hierfür verwendeten Bruttojahreslohn pro Beschäftigtem ist das Jahreszeitraummaterial der Bundesagentur für Arbeit (BA). Dieses enthält alle Meldungen der Arbeitgeber über sämtliche sozialversicherungspflichtige Beschäftigungsverhältnisse einschließlich der Bruttolöhne, die zunächst an die Krankenkassen und Rentenversicherungsträger und anschließend an die Bundesagentur für Arbeit übermittelt werden.¹ Erfasst werden somit alle sozialversicherungspflichtigen Bruttoentgelte, die innerhalb eines Jahres von allen sozialversicherungspflichtig beschäftigten Personen (ganzjährig oder mit Unterbrechungen) in einem oder mehreren Beschäftigungsverhältnissen, bei einem oder mehreren Betrieben, an einem oder mehreren Arbeitsorten verdient wurden. Nicht enthalten sind im Jahreszeitraummaterial Angaben über Beamte, Selbständige und mithelfende Familienangehörige, dafür aber sozialversicherungspflichtige Zahlungen an die Arbeitnehmer für Überstunden, Urlaubsgeld, Weihnachtsgeld oder andere Lohnkomponenten wie 13. Monatsgehalt oder Lohnfortzahlungen im Krankheitsfall.

Die Auswertung des Jahreszeitraummaterials des Jahres 2003 umfasst alle Meldungen der Arbeitgeber nach 18 Monaten, d. h. bis zum 30.06.2005, und damit über 99 % aller sozialversicherungspflichtigen Beschäftigungs-

¹ Grundlage für die Erfassung und Übermittlung der Daten an die Sozialversicherungsträger war bis Ende 1998 die Datenübermittlungsverordnung (DÜVO) und seit 1999 die Datenerfassungs- und -übermittlungsverordnung (DEÜV).

verhältnisse.² Erfasst werden 38,2 Mio. Beschäftigungsverhältnisse von rd. 30 Mio. Arbeitnehmern, die irgendwann einmal im Laufe des Jahres 2003 eine, in manchen Fällen auch nur eine kurzfristige sozialversicherungspflichtige Beschäftigung hatten.³ Diese Zahl ist höher als die Zahl der Personen zu einem Stichtag oder im Jahresdurchschnitt und vermittelt somit ein vollständigeres Bild aller Bruttolöhne und Bruttogehälter in sozialversicherungspflichtiger Beschäftigung.

Gerade strukturschwache Regionen sind dadurch gekennzeichnet, dass ein größerer Teil der Arbeitnehmer nicht das ganze Jahr über beschäftigt ist und dass oft in befristeten oder nicht dauerhaften Beschäftigungsverhältnissen weniger pro Tag und Stunde verdient wird als in ganzjähriger Beschäftigung. Auch bei normaler Konjunktur sind etwa nur drei Viertel aller Arbeitnehmer ganzjährig beschäftigt. Mit der Erfassung von ganzjährigen und auch kurzfristigen Beschäftigungsverhältnissen wird die tatsächliche regionale Einkommenssituation in sozialversicherungspflichtiger Beschäftigung umfassender abgebildet. Mit dem Jahreszeitraummaterial kann jedes einzelne Beschäftigungsverhältnis, und damit auch die Summe aller Verdienste, eindeutig einem Betrieb und damit einer Region (gemessen am Arbeitsort) zugewiesen werden. Ebenso ist eine Auswertung des Jahreszeitraummaterials am Wohnort der sozialversicherungspflichtig Beschäftigten möglich.

Die Bruttolohn- und Gehaltssumme⁴ $BLS_i(t)$ der Region i und des Jahres t lässt sich aus der Zahl der Beschäftigungsfälle $BF_i(t)$, der durchschnittlichen Dauer aller sozialversicherungspflichtigen Beschäftigungsverhältnisse $TpBF_i(t)$ (Tage pro Beschäftigungsfall) – und damit aus dem Beschäfti-

² Dieses Material ist noch nicht ganz vollständig wie die Vollerfassung nach 36 Monaten. Im Jahr 2001 führten die fehlenden Meldungen (zwischen der Vollerfassung und der Erfassung nach 18 Monaten) beim Bruttojahreslohn pro Beschäftigtem auf Kreisebene zu einer Untererfassung von 0,2 % bis 2,6 % und lagen im regionalen Durchschnitt bei rd. 0,6 %.

³ Hierbei wurden 12,2 Mio. Beschäftigungsverhältnisse von rd. 7,8 Mio. geringfügig Beschäftigten (davon ungefähr ein Drittel im Nebenerwerb und zwei Drittel im Haupterwerb – siehe u. a. Arntz et al. 2003) nicht berücksichtigt. Eine Einbeziehung ist ohne weiteres möglich, jedoch spielen die geringfügigen Beschäftigungsverhältnisse im weiteren Vorgehen der Lohnsummenkorrektur keine Rolle.

⁴ Genauer: Bruttolohnsumme und Bruttogehaltssumme.

gungsvolumen $BV_i(t)$ – und dem durchschnittlichen regionalen Bruttotageslohn $LpT_i(t)$ mit der folgenden Formel ermitteln:

$$BLS_i(t) = BV_i(t) \cdot LpT_i(t) = BF_i(t) \cdot TpBF_i(t) \cdot LpT_i(t)$$

Dividiert man diese gemessene Bruttolohn- und Gehaltssumme $BLS_i(t)$ durch die Anzahl der Beschäftigten $B_i(t)$, so erhält man den durchschnittlichen regional gemessenen Bruttojahreslohn pro Beschäftigtem $LpB_i(t)$. Für die regionalen Auswertungen werden die Beschäftigten und die Bruttolohn- und Gehaltssummen an dem Arbeitsort ausgewiesen, an dem das Hauptbeschäftigungsverhältnis besteht. Dieses ist für eine Person mit mehreren Beschäftigungsverhältnissen das mit der längeren Dauer bzw. bei gleichlangen Beschäftigungsverhältnissen das mit dem höheren Entgelt. Damit können alle ganzjährigen oder mehrfachen Beschäftigungsverhältnisse mit den dazugehörigen Arbeitsentgelten der richtigen Region zugeordnet werden und die obige Gleichung kann wie folgt erweitert werden:

$$BLS_i(t) = B_i(t) \cdot JpB_i(t) \cdot TpBF_i(t) \cdot LpT_i(t) = B_i(t) \cdot LpB_i(t)$$

Dabei drückt $JpB_i(t)$ die durchschnittliche Anzahl von Jobs pro Beschäftigtem in der Region i im Jahr t aus.

1.2 Beitragsbemessungsgrenze der sozialversicherungspflichtigen Einkommen

Im Jahreszeitraummaterial werden Einkommensangaben – bedingt durch das Meldeverfahren an die Sozialversicherungsträger – nur bis zur jeweils gültigen Beitragsbemessungsgrenze⁵ (BBG) erfasst. Alle Einkommen, die über der Beitragsbemessungsgrenze liegen, werden an der Beitragsbemessungsgrenze abgeschnitten und die Fälle dort kumuliert. Der in einer Region gemessene Mittelwert aller Beschäftigungsverhältnisse liegt somit niedriger als der „tatsächliche“ Mittelwert. Die Einkommen werden jedoch in voller Höhe erfasst, wenn ein Arbeitnehmer innerhalb eines Jahres mehrere Beschäftigungsverhältnisse hat und zwar die insgesamt erzielte Lohnsumme über der Beitragsbemessungsgrenze liegt, die einzelnen Bruttoverdienste aber geringer ausfallen.

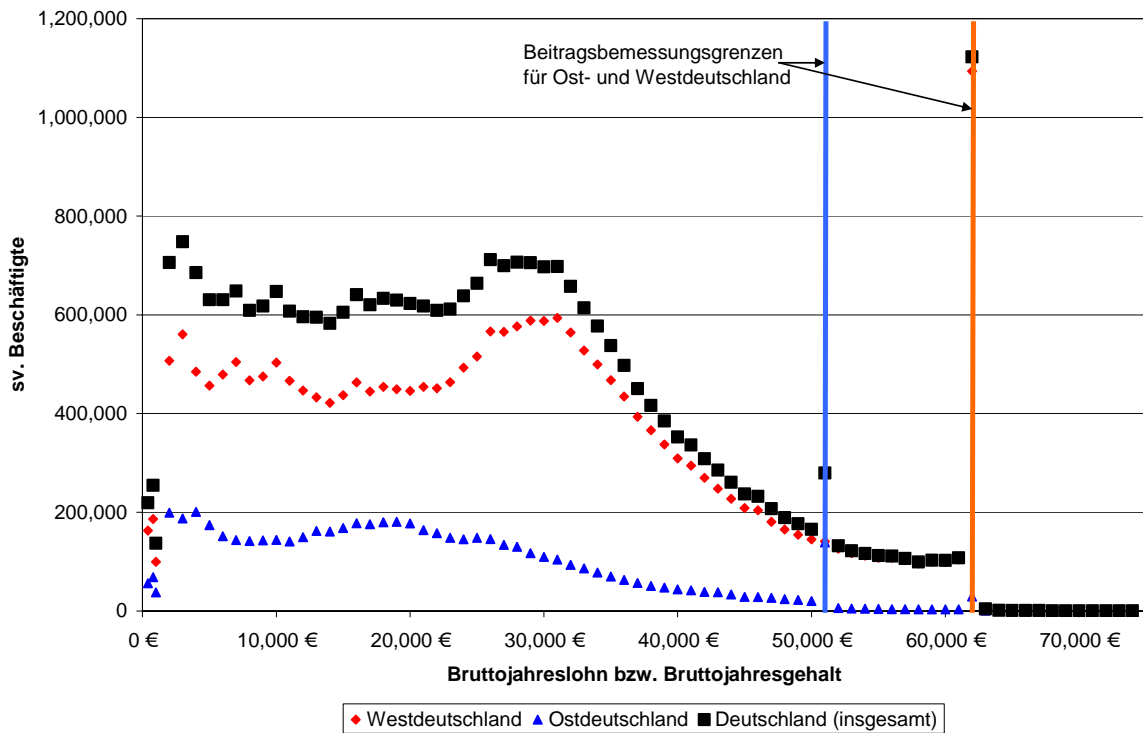
⁵ Die Beitragsbemessungsgrenze lag im Jahr 2003 in Westdeutschland bei 61.200 € und in Ostdeutschland bei 51.000 €.

Abbildung 1 zeigt die gemessene zensierte Einkommensverteilung für West-, Ost- und Gesamtdeutschland. Gut zu erkennen ist das deutlich niedrigere Niveau der Einkommensverteilung in Ostdeutschland im Vergleich zu Westdeutschland. Dieses wird besonders offensichtlich in Tabelle 1, in der die Anteile von Beschäftigten mit Einkommen oberhalb der Beitragsbemessungsgrenze für die zehn stärksten und drei schwächsten west- und ostdeutschen Kreise für das Jahr 2003 dargelegt werden. Der regionale Anteil der „Gutverdiener“ an allen Beschäftigten bewegt sich in den westdeutschen Kreisen zwischen 0,8 % und 18,7 % und in den ostdeutschen Kreisen trotz einer um rd. 10.000 € niedrigeren Beitragsbemessungsgrenze nur zwischen 0,8 % und 6,6 %. Der Anteil dieser Personengruppe mit einem Einkommen über der Beitragsbemessungsgrenze liegt bei vollzeitbeschäftigten Männern in jedem Kreis sehr viel höher und erreicht in Erlangen sogar rd. 32 %. Die regionalen Verdienstmöglichkeiten im Hochlohnbereich streuen sowohl in den Regionen als auch zwischen Männern und Frauen in Voll- und Teilzeitbeschäftigungen zum Teil erheblich.⁶

Die Analyse von einseitig zensierten Einkommensverteilungen erfolgt häufig nur anhand der Größen Median und Quartil, da diese meist nicht - im Gegensatz zur Varianz und zum Mittelwert - von der Einkommensverteilung oberhalb der Beitragsbemessungsgrenze beeinflusst werden (siehe z. B. Lankes 1997). Auf Basis von Individualdaten kann mit Tobit- und Probitmodellen oder Quartilsregressionen die Abhängigkeit der Einkommensverteilung von regionalen, arbeitsmarktrelevanten und sozioökonomischen Variablen analysiert werden (Merz 2002; Fitzenberger/Reize 2002). Allerdings können diese Modelle keine Aussagen zur Einkommenshöhe jenseits der Beitragsbemessungsgrenze treffen. Einen Ausweg bietet das Verfahren der multiplen Imputation an, dem eine Markov-Ketten-Monte-Carlo-Technik zu Grunde liegt (Gartner, Rässler 2005).

⁶ Auch zwischen Arbeitern und Angestellten gibt es erhebliche Differenzen: Während 1995 fast jeder vierte männliche Angestellte in Westdeutschland ein Entgelt über der Beitragsbemessungsgrenze bezogen hat, ist es gerade einmal nur jede 1.666te Arbeiterin in Ostdeutschland. Weiterhin ist die Einkommensverteilung stark altersabhängig. Während die Entgelte von jüngeren Beschäftigten noch deutlich ansteigen, stagnieren sie bei älteren Arbeitnehmern. Den Höhepunkt erreichen die Entgelte bei den 50- bis 54-Jährigen und verharren dann auf diesem Niveau. (Lankes 1997)

Abbildung 1: Gemessene Einkommensverteilung der sv. Beschäftigten in West-, Ost- und Gesamtdeutschland 2003



Quelle: Jahreszeitraummaterial der Bundesagentur für Arbeit 2003

Tabelle 1: Anteil der sv. Beschäftigten bzw. der vollzeitbeschäftigten Männer mit einem Bruttोजahreslohn über der BBG an allen sv. Beschäftigten im Jahr 2003 für die zehn stärksten und die drei schwächsten west- und ostdeutschen Kreise

Westdeutsche Kreise	Anteil der	
	sv. Beschäftigten	Männer in Vollzeit
Erlangen, Stadt	18,7%	31,9%
Main-Taunus-Kreis	14,3%	22,6%
München	13,9%	22,3%
Frankfurt am Main	12,7%	20,9%
München, Stadt	12,0%	21,6%
Groß-Gerau	10,3%	15,4%
Stuttgart, Stadt	10,2%	18,7%
Böblingen	9,9%	16,1%
Hochtaunuskreis	9,9%	17,6%
Wolfsburg, Stadt	9,7%	14,1%
⋮	⋮	⋮
Regen	1,0%	1,7%
Freyung-Grafenau	0,9%	1,5%
Wittmund	0,8%	1,4%
Ø Westdeutschland	4,7%	8,4%

Ostdeutsche Kreise	Anteil der	
	sv. Beschäftigten	Männer in Vollzeit
Berlin-Ost	6,6%	10,6%
Dresden, Stadt	5,7%	9,4%
Jena, Stadt	5,5%	9,6%
Potsdam, Stadt	5,1%	8,3%
Leipzig, Stadt	5,0%	8,4%
Schwerin	4,0%	6,6%
Erfurt, Stadt	3,9%	6,4%
Oberhavel	3,8%	5,9%
Rostock	3,8%	5,9%
Teltow-Fläming	3,8%	5,9%
⋮	⋮	⋮
Eichsfeld	0,9%	1,5%
Loebau-Zittau	0,9%	1,5%
Mittlerer Erzgebirgs-kreis	0,8%	1,3%
Ø Ostdeutschland	2,8%	4,7%

Im vorliegenden Fall der regionalen Lohnkorrektur kann allerdings nicht auf die oben beschriebenen Verfahren zurückgegriffen werden, da es sich hierbei nicht um die Analyse von Individualdaten, sondern von regional aggregierten Daten handelt. Um die realen Verdienstmöglichkeiten in den einzelnen Regionen dennoch möglichst exakt abzubilden, muss der durch das institutionelle Abschneideverfahren bedingte Fehler bereinigt werden. Dazu ist es notwendig, die Einkommensverteilungen über der Beitragsbemessungsgrenze sowie die „abgeschnittenen“ Bruttolohn- und Gehaltssummen hinzuschätzen und die Mittelwerte **getrennt nach Männern und Frauen in Voll- und Teilzeitbeschäftigungen** sowie für **jede Region einzeln** zu korrigieren.

2 Korrekturverfahren

2.1 Allgemeines Vorgehen

Da die tatsächlichen Verteilungen oberhalb der Beitragsbemessungsgrenze nicht bekannt sind, wird **für jede Region** eine **hypothetische Einkommensverteilungskurve** gesucht, bei der die letzten „Hochlohnklassen“ vor der Abschneidegrenze mit der gemessenen Kurve übereinstimmen. Diese hypothetische Verteilungsfunktion wird dann über die Abschneidegrenze hinaus „verlängert“ – und zwar unter der **Annahme der Log-Normalverteilung**. Diese Verteilungsfunktion wird in der aktuellen Literatur am häufigsten für die Analyse und Approximation von Lohn- und Einkommensverteilungen herangezogen, da die beobachtete Einkommensverteilung in vielen Ländern linkssteil verläuft (Blümle 1975; Neal, Rosen 2000; Limpert et al. 2001)⁷. Bereits 1950 lieferte Roy einen viel zitierten stochastischen Erklärungsansatz, mit dem er zeigt und empirisch belegt, dass das multiplikative Zusammenwirken unabhängiger und gleichverteilter individueller Grundfähigkeiten (Faktoren) lognormalverteilte Arbeitsproduktivitäten und damit auch lognormalverteilte Einkommen impliziert (Sheldon 2002). Da Lohnerhöhungen in der Regel prozentual erfolgen und damit hohe Einkommen stärker angehoben werden als niedrigere, verstärken sie eine rechtsschiefe (linkssteile) Verteilung und damit auch die Lohnspreizung.

⁷ Neben der Log-Normalverteilung werden zur Abbildung von Einkommensverteilungen auch die Interpolationsfunktionen Paretoverteilung und verallgemeinerte Beta-Verteilung der zweiten Art angewendet (Müller 2005: 12).

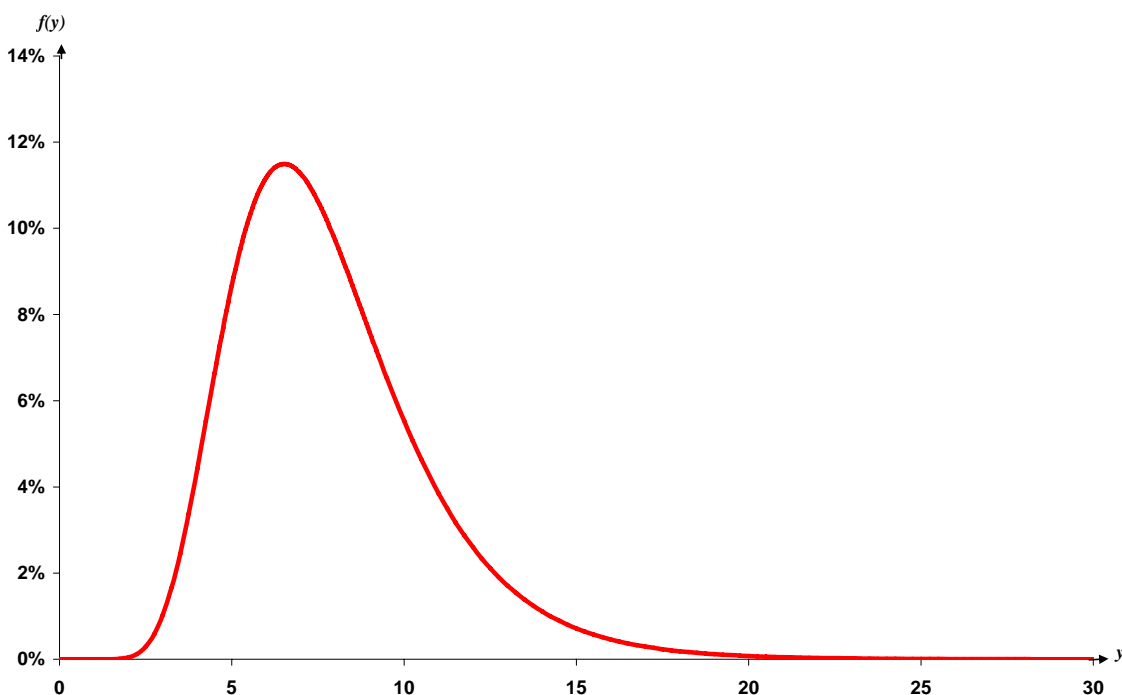
Mit dem hier vorgestellten Verfahren wird die Zahl der Personen, die in der Einkommensklasse an der Beitragsbemessungsgrenze erfasst wird, mit der Log-Normalverteilung so oberhalb der Beitragsbemessungsgrenze verteilt, dass die Zahl aller Personen mit einem Einkommen über der Beitragsbemessungsgrenze nach dieser Korrektur mit der Zahl der Personen an der Beitragsbemessungsgrenze übereinstimmt. Damit erreicht man eine realistische Korrektur der regionalen Mittelwerte nach oben. Der Median und die Quartile der Lohnverteilung bleiben durch die Lohnkorrektur unberührt, da in fast keinem Kreis mehr als 25 % aller Personen ein Einkommen über der Beitragsbemessungsgrenze beziehen.

2.2 Eigenschaften der Log-Normalverteilung

Die zur Approximation der regionalen Einkommensverteilungen verwendete Log-Normalverteilung ist linkssteil (rechtsschief) und in Abbildung 2 dargestellt. Bei linkssteilen Verteilungen ist der Median kleiner als der arithmetische Mittelwert der Verteilung. Je größer die Differenz zwischen Erwartungswert und Median ist, desto ausgeprägter fällt im Allgemeinen die Schiefe der Verteilung aus. Die Dichtefunktion der (unnormierten) Log-Normalverteilung hat die folgende mathematische Form:

$$f_L(y) = \frac{1}{\sigma\sqrt{2\pi}} \frac{\chi}{y} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) \quad \text{mit} \quad \int_0^{\infty} f_L(y) dy = \chi, \quad \mu \in \mathbb{R} \quad \text{und} \quad \sigma \in \mathbb{R}^+$$

Abbildung 2: Darstellung der Log-Normalverteilung



Der **Erwartungswert** und die **Varianz** der Log-Normalverteilung berechnen sich über

$$E(y) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad \text{bzw.} \quad E(\ln(y)) = \mu$$

$$Var(y) = \exp(2\mu + \sigma^2)(\exp\sigma^2 - 1) \quad \text{bzw.} \quad Var(\ln(y)) = \sigma^2$$

Mit der Transformation $y \rightarrow \ln(y)$ geht die Log-Normalverteilung in die Normalverteilung über. Die Parameter μ , σ und χ der Log-Normalverteilung werden auch als Skalen- (μ), Form- (σ) und Flächen- bzw. Normierungsparameter (χ) interpretiert. Der Flächen- bzw. Normierungsparameter χ beeinflusst weder den Erwartungswert noch die Varianz der Log-Normalverteilung, sondern nur die Fläche unterhalb der Verteilungsfunktion.

2.3 Anpassung der Log-Normalverteilung an die Einkommensverteilung

Oberhalb der Beitragsbemessungsgrenze ($i > BBG$) soll die Einkommensverteilung $P(EK_{ik})$ für jede Region k durch die diskrete Dichtefunktion der Log-Normalverteilung

$$P(EK_{ik}) = \frac{\chi_k}{\sigma_k \sqrt{2\pi}} \frac{1}{EK_{ik}} \exp\left(-\frac{(\ln(EK_{ik}) - \mu_k)^2}{2\sigma_k^2}\right)$$

beschrieben werden. Dabei gibt $P(EK_{ik})$ die Anzahl der Personen mit Einkommen EK_{ik} aus der Einkommensklasse i in der Region k an. Die drei regionsspezifischen Parameter – der Skalenparameter μ_k , der Formparameter σ_k und der Flächenparameter χ_k – werden über das Least-Square-Verfahren geschätzt. Bei der Anpassung der analytischen Verteilungsfunktion an die empirisch gemessene Einkommensverteilung ergeben sich in der Regel erhebliche Abweichungen, insbesondere in den unteren Einkommensbereichen. Aus diesem Grund werden die Parameter nur über die letzten n Einkommensklassen (1.000€-Klassen) der gemessenen Einkommensverteilung $\overline{P(EK_{ik})}$ vor der Beitragsbemessungsgrenze (BBG) ermittelt. Somit muss die folgende Funktion minimiert werden:

$$LS = \underset{\chi, \mu, \sigma}{MIN} \left(\sum_{i=BBG-n}^{BBG-1} \left(\overline{P(EK_{ik})} - P(EK_{ik}) \right)^2 \right)$$

$$= \underset{\chi, \mu, \sigma}{MIN} \left(\sum_{i=BBG-n}^{BBG-1} \left(\overline{P(EK_{ik})} - \frac{\chi_k}{\sigma_k \sqrt{2\pi}} \frac{1}{EK_{ik}} \exp\left(-\frac{(\ln(EK_{ik}) - \mu_k)^2}{2\sigma_k^2}\right) \right)^2 \right)$$

Das zur Minimierung der Zielfunktion oben verwendete Least-Square-Verfahren eignet sich besonders für die Anpassung der Parameter, da insbesondere große Abweichungen der geschätzten von der gemessenen Einkommensverteilung vermieden werden sollen. Für die angepasste Einkommensverteilung über der Beitragsbemessungsgrenze sollten jedoch noch die folgenden beiden Punkte erfüllt werden:

1. Die an der Beitragsbemessungsgrenze erfasste Zahl der Beschäftigten aus der gemessenen Einkommensverteilung $\overline{P(EK_{(BBG)k})}$ soll der Zahl der Beschäftigten nach der Schätzung der Einkommensverteilung oberhalb der Beitragsbemessungsgrenze entsprechen (starke Nebenbedingung), d. h.

$$\overline{P(EK_{(BBG)k})} = \sum_{i=BBG}^{\infty} P(EK_{ik}) = \chi_k \cdot \left(\sum_{i=BBG}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{EK_{ik}} \exp\left(-\frac{(\ln(EK_{ik}) - \mu_k)^2}{2\sigma_k^2}\right) \right)$$

2. Die geschätzte Einkommensverteilung über der Beitragsbemessungsgrenze soll die gemessene Einkommensverteilung annähernd stetig verlängern. Dies setzt voraus, dass die Einkommensklassen vor und oberhalb der Beitragsbemessungsgrenze aus der gemessenen und der geschätzten Verteilung annähernd übereinstimmen (schwache Nebenbedingung), d. h.:

$$\overline{P(EK_{(BBG-1)k})} \approx P(EK_{(BBG-1)k}) = \frac{\chi_k}{\sigma_k\sqrt{2\pi}} \frac{1}{EK_{(BBG-1)k}} \exp\left(-\frac{(\ln(EK_{(BBG-1)k}) - \mu_k)^2}{2\sigma_k^2}\right)$$

Die erste Nebenbedingung lässt sich in die Funktion direkt einbinden, indem man die Nebenbedingung nach χ_k auflöst und entsprechend einsetzt. Die zweite Nebenbedingung kann über einen Lagrange-Parameter λ berücksichtigt werden. Man erhält damit die Gleichung:

$$LS = \underset{\chi, \mu, \sigma}{MIN} \left[\begin{aligned} & \frac{1}{n} \sum_{i=BBG-n}^{BBG-1} \left(\overline{P(EK_{ik})} - \frac{\chi}{\sigma\sqrt{2\pi}} \frac{1}{EK_{ik}} \exp\left(-\frac{(\ln(EK_{ik}) - \mu)^2}{2\sigma^2}\right) \right)^2 \\ & + \lambda \left(\overline{P(EK_{(BBG-1)k})} - \frac{\chi}{\sigma\sqrt{2\pi}} \frac{1}{EK_{(BBG-1)k}} \exp\left(-\frac{(\ln(EK_{(BBG-1)k}) - \mu)^2}{2\sigma^2}\right) \right)^2 \end{aligned} \right] \text{ mit}$$

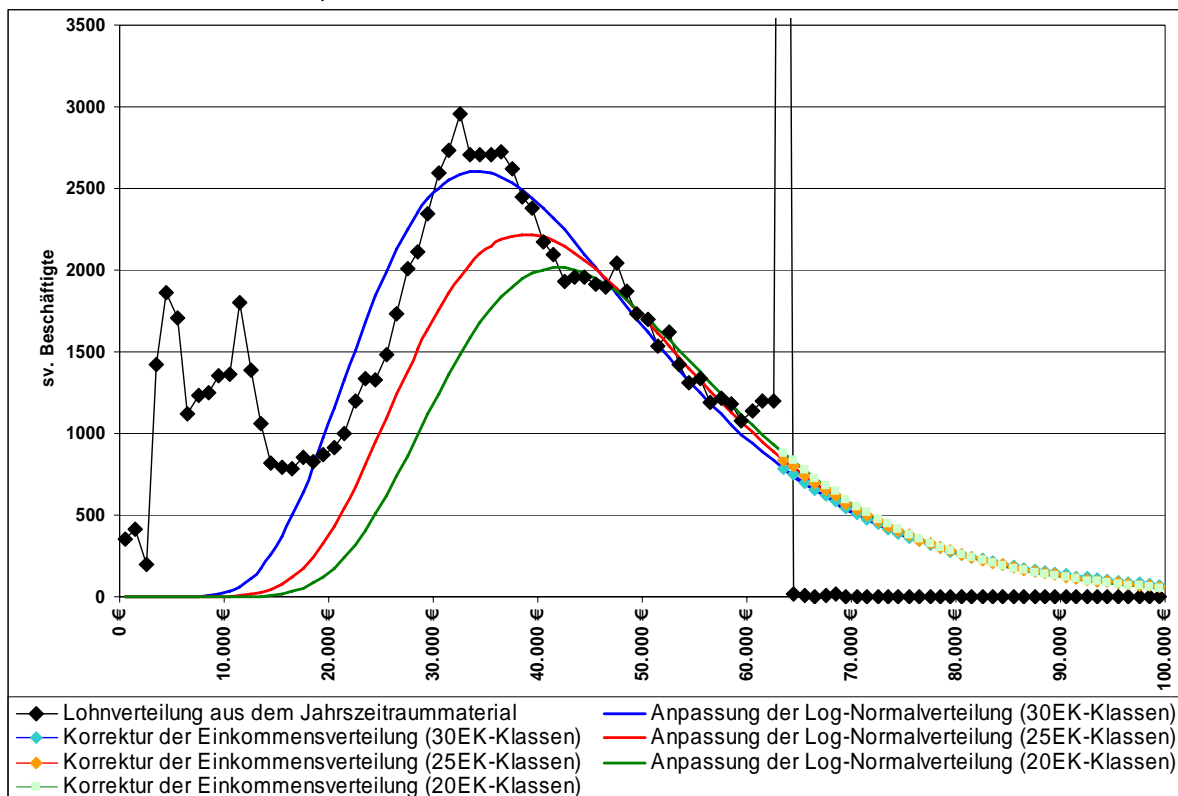
$$\chi = \frac{\overline{P(EK_{(BBG)k})}}{1 - \sum_{i=BBG}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{EK_{ki}} \exp\left(-\frac{(\ln(EK_{ki}) - \mu)^2}{2\sigma^2}\right)}$$

Diese Minimierungsfunktion hat für jede Region drei Parameter und **eine starke** Nebenbedingung. Infolgedessen besitzt sie nur zwei Freiheitsgra-

de. Die Schätzung der regionalen Parameter μ_k , σ_k und χ_k der Log-Normalverteilung erfolgt für jeden Stadt- und Landkreis einzeln. Es bleibt nur noch zu bestimmen, wie stark die Nebenbedingung 2 in der Gleichung berücksichtigt wird, und wie viele Einkommensklassen n zur Bestimmung der Parameter μ_k , σ_k und χ_k herangezogen werden.

Die Abbildung 3 zeigt die Einkommensverteilung von vollzeitbeschäftigten Männern im Landkreis Esslingen für $n=20$, $n=25$ und $n=30$ ohne Berücksichtigung der Nebenbedingung 2 ($\lambda=0$). Es zeigt sich, dass sich die Verteilungen oberhalb der Beitragsbemessungsgrenze kaum voneinander unterscheiden. Ohne Berücksichtigung der Nebenbedingung 2 würde sich allerdings die Zahl der Beschäftigten mit Einkommen an der Beitragsbemessungsgrenze gegenüber der Zahl der Beschäftigten in der Einkommensklasse vor der Beitragsbemessungsgrenze nahezu halbieren, da bei fast allen regionalen Einkommensverteilungen in den letzten zwei oder drei Einkommensklassen vor der Beitragsbemessungsgrenze die Zahl der Personen konstant bzw. ansteigend ist.

Abbildung 3: Unterschiedliche Anpassungen der Einkommensverteilung von vollzeitbeschäftigten Männern in Esslingen 2003 für $\lambda=0$ und $n=20$, $n=25$ und $n=30$



In den Abbildungen 4a und 4b wurde für $n=20$ bzw. $n=30$ der Lagrange-Parameter λ variiert. Dabei ist es nicht notwendig, die 2. Nebenbedingung exakt zu erfüllen ($\lambda=\infty$), da sich dann die Log-Normalverteilung kaum an die gegebene Einkommensverteilung anpasst. Es ist offensichtlich, dass der Sprung an der Beitragsbemessungsgrenze für $n=30$ und $\lambda=1$ gegenüber $n=20$ und $\lambda=1$ sehr viel stärker ist. Bei Verwendung von $n=30$ Einkommensklassen zur Bestimmung der Parameter fällt die korrigierte Einkommensverteilung über der Beitragsbemessungsgrenze folglich flacher aus als bei $n=20$.

Abbildung 4a: Unterschiedliche Anpassungen der Einkommensverteilung von vollzeitbeschäftigten Männern in Esslingen 2003 für $\lambda=0$ und $\lambda=1$ und $n=20$

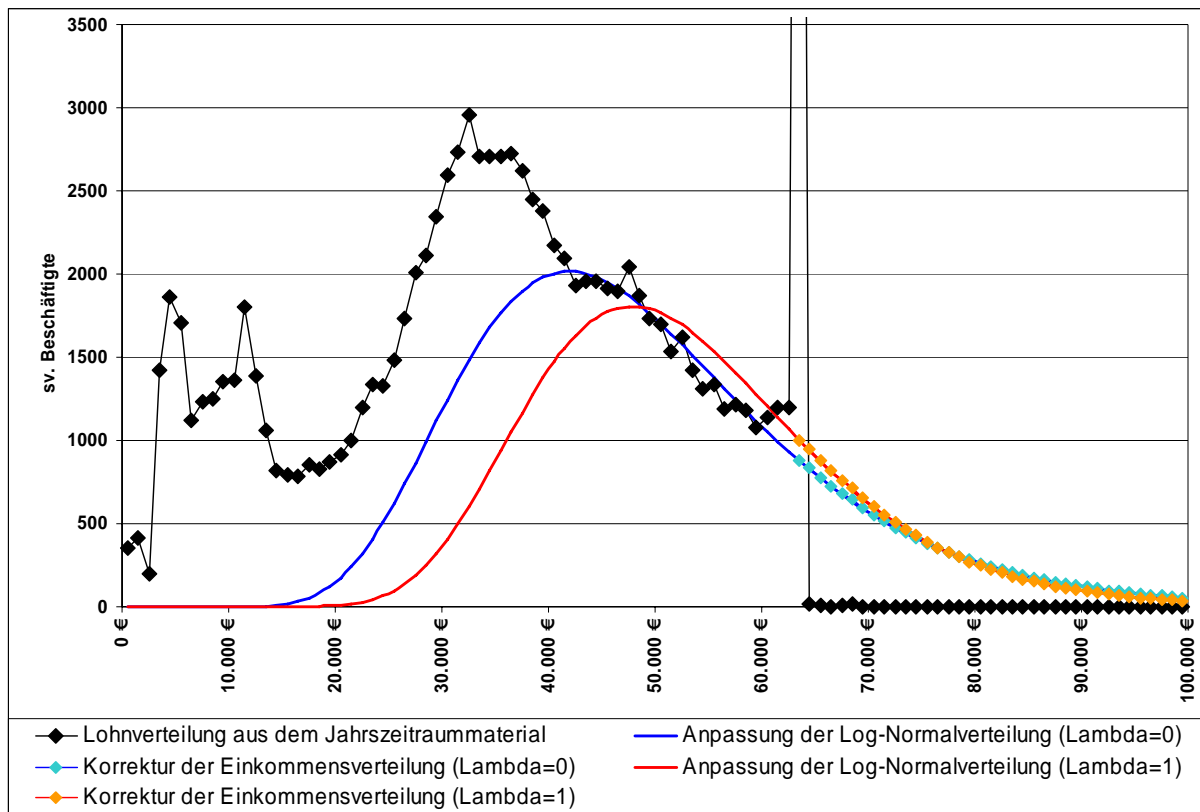
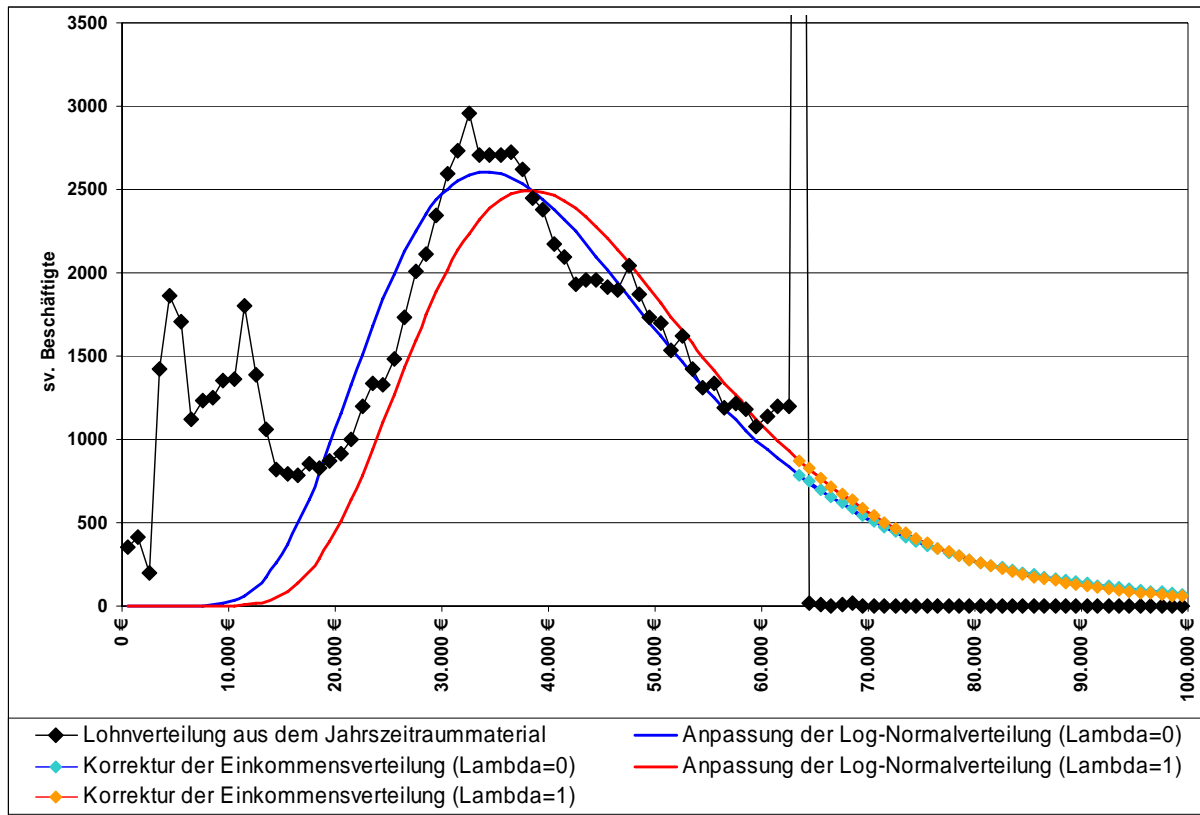


Abbildung 4b: Unterschiedliche Anpassungen der Einkommensverteilung von vollzeitbeschäftigten Männern in Esslingen 2003 für $\lambda=0$ und $\lambda=1$ und $n=30$



Weiterhin ist anzumerken, dass die Einkommensverteilung nicht in allen Kreisen ein eindeutiges Maximum aufweist. Die Einkommensverteilung in ostdeutschen Kreisen zeigt häufig einen monoton fallenden Verlauf ohne jegliches Maximum. Ebenso trifft dies für die Einkommensverteilung von Teilzeitbeschäftigten zu. Grundsätzlich fallen die regionalen Einkommensverteilungen der Einkommensklassen bis 25.000 € für einzelne Kreise bzw. für Frauen und Männer wie auch für Voll- und Teilzeitbeschäftigte sehr unterschiedlich aus. Daher wurden für jeden Kreis die Einkommensverteilungen getrennt nach vollzeitbeschäftigten Männern und Frauen sowie für Teilzeitbeschäftigte insgesamt korrigiert. Die Verteilungen über der Beitragsbemessungsgrenze sind unbekannt. Aufgrund der monoton fallenden Eigenschaft der Log-Normalverteilung nach ihrem Maximum sowie der starken und schwachen Nebenbedingung konvergiert die geschätzte Lohnverteilung bei Einkommensklassen zwischen 100.000 € und 150.000 € je nach Kreis gegen Null. Da jedoch z.B. Führungskräfte häufig mehr als 150.000 € pro Jahr verdienen, werden die Einkommen über der Beitragsbemessungsgrenze mit der hier beschriebenen Lohnsummenkor-

rektur vermutlich noch leicht unterschätzt. Die Annahme einer stetigen Verteilung direkt oberhalb der Beitragsbemessungsgrenze ist sicherlich für die folgenden 20 bis 40 Einkommensklassen gerechtfertigt. Dahinter sind die kreisspezifischen Einkommensklassen nur noch gering besetzt. Eine fortführende stetige Verteilungsannahme kann dann nicht mehr begründet werden. Eine kreisspezifische Annäherung der unteren und obersten Einkommensklassen ist mit keiner analytischen Verteilungsfunktion zu leisten (Krupp 1968, Müller 2005: 12).

3 Weitere Ergebnisse des Korrekturverfahrens

Wichtig ist bei diesem Korrekturverfahren, dass die Verteilungskurve bis zur west- bzw. der ostdeutschen Abschneidegrenze die tatsächlichen Messwerte wiedergibt. Lediglich der Verlauf oberhalb der Beitragsbemessungsgrenze, d. h. für die Hochlohngruppen, ist geschätzt. Damit erreicht man eine Korrektur der regionalen Mittelwerte nach oben. Das Ausmaß der Korrektur für die einzelnen Kreise wird vor allem durch die Anzahl der Personen und durch die Steigung der angepassten Einkommensverteilung an der Beitragsbemessungsgrenze bestimmt. Der Median und zumeist auch die Quartile der Lohnverteilung bleiben von der Lohnkorrektur unberührt.

Für den korrigierten Bruttojahreslohn pro Beschäftigtem – der im Rahmen des IAB-Forschungsprojekts zur Neuabgrenzung der Fördergebiete berücksichtigt wird – wurden nur die letzten 20 Einkommensklassen zur Anpassung der Parameter μ , σ und χ der Log-Normalverteilung verwendet und die zweite Nebenbedingung (Stetigkeits-Bedingung) mit $\lambda=1/2$ berücksichtigt. Dass nur 20 Einkommensklassen verwendet wurden, hat u. a. den Hintergrund, dass die Anpassung im fallenden Bereich der Einkommensverteilung erfolgen sollte, und dies im Falle der ostdeutschen Kreise für $n=30$ nicht immer gegeben ist. Außerdem kann für $n=20$ zusätzlich die Stetigkeitsbedingung (schwache Nebenbedingung) eher erfüllt werden.

Die Auswirkungen des Korrekturverfahrens mit $n=20$ Einkommensklassen ohne bzw. mit Berücksichtigung der schwachen Nebenbedingung ($\lambda=0$ bzw. $\lambda=1/2$) auf die zehn einkommensstärksten Kreise in Westdeutschland sind in Tabelle 2 und in Ostdeutschland in Tabelle 3 dargestellt.

Tabelle 2: Auswirkungen des Korrekturverfahrens (mit $n=20$) auf die zehn einkommensstärksten westdeutschen Kreise

	Anteil der Beschäftigten bzw. der Bruttolohnsumme oberhalb der BBG		Anteil der korrigierten Bruttolohnsumme oberhalb der BBG		Erhöhung der gemessenen Bruttolohnsumme	
	Beschäftigte	Bruttolohnsumme (gemessen)	$\lambda=0$	$\lambda=1/2$	$\lambda=0$	$\lambda=1/2$
Erlangen, Stadt	18,7%	35,0%	43,0%	41,9%	19,5%	16,4%
Main-Taunus-Kreis	14,3%	26,9%	33,2%	32,6%	8,6%	7,4%
München	13,9%	25,7%	32,5%	28,9%	9,2%	6,9%
Frankfurt, Stadt	12,7%	23,6%	28,9%	28,2%	7,2%	5,9%
München, Stadt	12,0%	23,4%	29,7%	31,6%	8,4%	7,6%
Groß-Gerau	10,3%	20,3%	27,2%	27,1%	9,5%	9,3%
Stuttgart, Stadt	10,2%	19,6%	24,6%	24,2%	5,0%	4,3%
Böblingen	9,9%	18,5%	25,0%	25,0%	7,0%	7,0%
Hochtaunuskreis	9,9%	20,1%	25,4%	24,7%	6,4%	5,2%
Wolfsburg, Stadt	9,7%	17,8%	24,0%	22,9%	7,0%	5,3%
Ø Westdeutschland	4,7%	11,1%	14,1%	13,8%	3,1%	2,7%

In Erlangen (Dresden) verdienen 18,7 % (5,7 %) aller Beschäftigten ein Einkommen oberhalb der Beitragsbemessungsgrenze von 61.200 € (51.000 €), was 35 % (13,3 %) der gemessenen Bruttolohnsumme des Kreises entspricht. Die Erhöhung der gemessenen Bruttolohn- und Gehaltssumme aufgrund des beschriebenen Verfahrens beträgt für $\lambda=0$ ($\lambda=1/2$) in Erlangen 19,5 % (16,4 %) und in Dresden 2,8 % (2,6 %). Die Lohnsummenkorrektur führt dazu, dass 18,7% der Beschäftigten aus Erlangen zu 43,0 % bzw. 41,9 % (statt 35 %) zur Bruttolohnsumme für $\lambda=0$ bzw. $\lambda=1/2$ beitragen. In Dresden erhöht sich der Anteil der Bruttolohnsumme der 5,7 % „gutverdienenden Beschäftigten“ an der korrigierten Bruttolohnsumme insgesamt von 13,3 % zu 14,8 % bzw. 14,7 % für $\lambda=0$ bzw. $\lambda=1/2$.

In Tabelle 4 wird deutlich, dass sich die Einkommensverteilungen von Männern und Frauen stark unterscheiden. Während in Westdeutschland 8,4% bzw. in Ostdeutschland 4,7 % aller vollzeitbeschäftigten Männer ein Einkommen über der Beitragsbemessungsgrenze von 61.200 € bzw. 51.000 € beziehen, trifft dies lediglich für 1,7 % bzw. 1,8 % der vollzeitbeschäftigten Frauen zu. Dies wirkt sich schließlich auch auf eine niedrigere Lohnsummenkorrektur aus.

Tabelle 3: Auswirkungen des Korrekturverfahrens (mit $n=20$) auf die zehn einkommensstärksten ostdeutschen Kreise

	Anteil der Beschäftigten bzw. der Bruttolohnsumme oberhalb der BBG		Anteil der korrigierten Bruttolohnsumme oberhalb der BBG		Erhöhung der gemessenen Bruttolohnsumme	
	Beschäftigte	Bruttolohnsumme (gemessen)	$\lambda=0$	$\lambda=1/2$	$\lambda=0$	$\lambda=1/2$
Berlin-Ost	6,6%	15,8%	17,3%	17,0%	3,1%	2,7%
Dresden, Stadt	5,7%	13,3%	14,8%	14,7%	2,8%	2,6%
Jena, Stadt	5,5%	12,8%	10,1%	14,0%	2,6%	2,4%
Potsdam, Stadt	5,1%	12,3%	13,4%	13,3%	2,2%	2,1%
Leipzig, Stadt	5,0%	12,2%	13,4%	13,3%	2,4%	2,2%
Schwerin, Landeshauptstadt	4,0%	9,8%	10,5%	10,5%	1,6%	1,5%
Erfurt, Stadt	3,9%	9,4%	10,1%	10,0%	1,5%	1,4%
Kreis Oberhavel	3,8%	10,0%	12,0%	11,7%	3,3%	2,8%
Rostock, Hansestadt	3,8%	9,4%	10,3%	10,2%	1,8%	1,7%
Kreis Teltow-Flaeming	3,8%	9,7%	11,8%	11,8%	3,2%	3,2%
Ø Ostdeutschland	2,8%	7,6%	8,8%	8,7%	1,6%	1,4%

Tabelle 4: Auswirkungen des Korrekturverfahrens (mit $n=20$) für Teilzeitbeschäftigte und vollzeitbeschäftigte Männer und Frauen in West-, Ost und Gesamtdeutschland

	Anteil der Beschäftigten bzw. der Bruttolohnsumme oberhalb der BBG		Anteil der korrigierten Bruttolohnsumme oberhalb der BBG		Erhöhung der gemessenen Bruttolohnsumme	
	Beschäftigte	Bruttolohnsumme (gemessen)	$\lambda=0$	$\lambda=1/2$	$\lambda=0$	$\lambda=1/2$
Ø Deutschland	4,3%	10,6%	13,3%	13,0%	2,9%	2,5%
Ø Deutschland Vollzeit Männer	7,7%	15,7%	18,8%	18,5%	4,3%	3,7%
Ø Deutschland Vollzeit Frauen	1,7%	4,5%	5,2%	5,2%	0,8%	0,7%
Ø Deutschland Teilzeit	0,3%	1,2%	1,3%	1,3%	0,1%	0,1%
Ø Westdeutschland	4,7%	11,1%	14,1%	13,8%	3,1%	2,7%
Ø Westdeutschland Vollzeit Männer	8,4%	16,3%	19,7%	19,2%	4,6%	3,9%
Ø Westdeutschland Vollzeit Frauen	1,7%	4,5%	5,3%	5,2%	0,8%	0,8%
Ø Westdeutschland Teilzeit	0,3%	1,3%	1,4%	1,4%	0,1%	0,1%
Ø Ostdeutschland	2,8%	7,6%	8,8%	8,7%	1,6%	1,4%
Ø Ostdeutschland Vollzeit Männer	4,7%	11,3%	13,5%	13,3%	2,6%	2,4%
Ø Ostdeutschland Vollzeit Frauen	1,8%	4,6%	5,1%	5,1%	0,6%	0,6%
Ø Ostdeutschland Teilzeit	0,2%	0,8%	0,9%	0,9%	0,1%	0,1%

Das Ergebnis der Bereinigung für die Einkommensverteilung insgesamt (aufsummiert aus den Kreisergebnissen für vollzeitbeschäftigte Männer und Frauen und Teilzeitbeschäftigte) ist in Abbildung 5a für Westdeutschland und in Abbildung 5b für Ostdeutschland dargestellt. Die Abbildungen 6 bis 9 im Anhang veranschaulichen noch einmal die Ergebnisse des Korrekturverfahrens und die regionale Verteilung der einzelnen Größen.

Abbildung 5a: Gemessene und korrigierte Einkommensverteilung für Westdeutschland (ohne Berlin)

Gemessene und korrigierte Einkommensverteilung (Bruttojahreslohn pro Kopf)
für Westdeutschland 2003

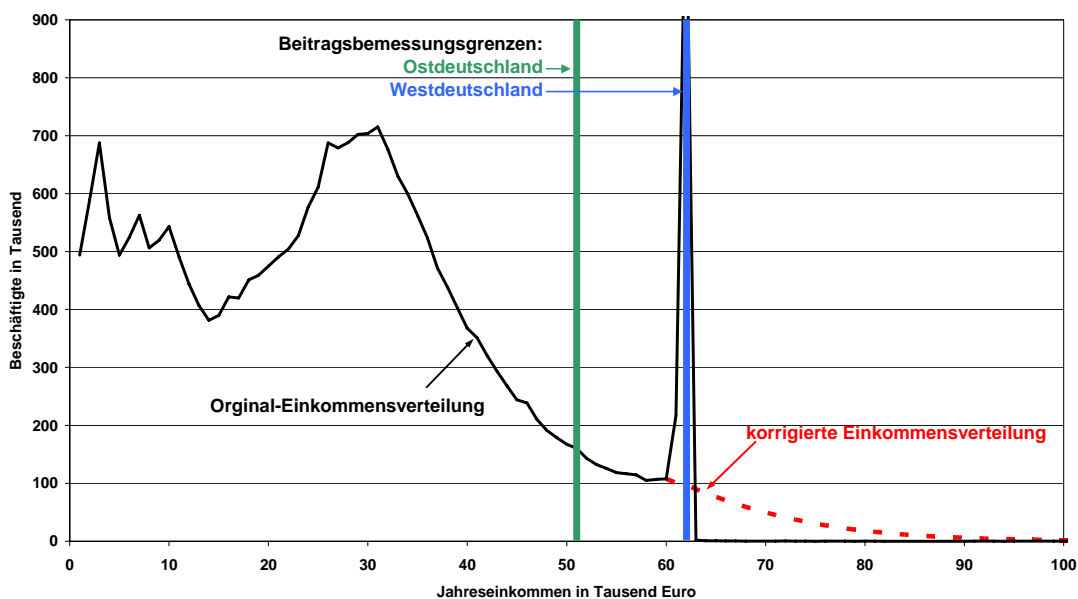
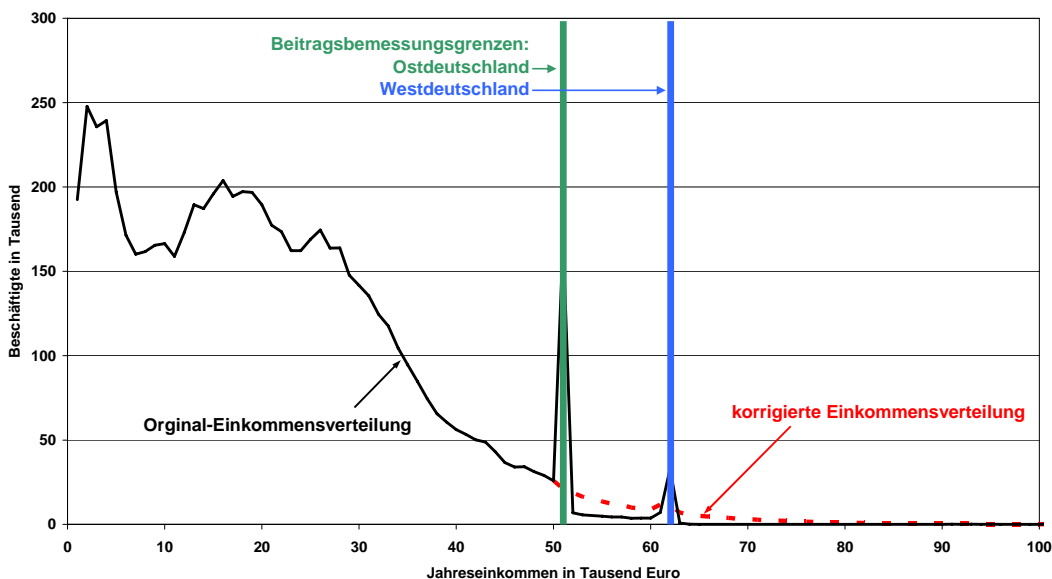


Abbildung 5b: Gemessene und korrigierte Einkommensverteilung für Ostdeutschland (mit Berlin)

Gemessene und korrigierte Einkommensverteilung (Bruttojahreslohn pro Kopf)
für Ostdeutschland 2003



Die korrigierten Bruttojahreslöhne pro Beschäftigten für das Jahr 2003 sind in Abbildung 6 im Anhang dargestellt. Gut zu erkennen ist die immer noch enorme Lohnspreizung zwischen West- und Ostdeutschland. Die höchsten Durchschnittslöhne werden in Baden-Württemberg, im Rhein-Main-Gebiet, im Ruhrgebiet, in und um München sowie in Hamburg verdient. Niedrige Bruttojahreslöhne pro Beschäftigten findet man hingegen nicht nur in Ostdeutschland, sondern ebenso in den Grenzregionen Bayerns und in ländlichen Gebieten von Niedersachsen, Schleswig-Holstein und Rheinland-Pfalz.

Abbildung 7 im Anhang zeigt die prozentuale Erhöhung (Korrektur) der gemessenen Bruttolohn- und Gehaltssumme für die einzelnen Kreise. Die stärksten Korrekturen fanden in Städten bzw. Ballungsräumen statt. Auch wenn es sich hierbei wieder vorwiegend um westdeutsche Regionen handelt, so stechen neben einigen ostdeutschen Städten vor allem die an Berlin angrenzenden Brandenburger Kreise Oberhavel und Teltow-Fläming in der Karte hervor.

Ein ähnliches Bild zeigen auch die Abbildungen 8 und 9 im Anhang, in denen zum einen der Anteil der Beschäftigten mit einem Einkommen oberhalb der Beitragsbemessungsgrenze abgebildet wird und zum anderen der Anteil der korrigierten Bruttolohnsumme oberhalb der Beitragsbemessungsgrenze insgesamt.

4 Fazit

Aufgrund des institutionellen Abschneideverfahrens ist die tatsächliche regionale Verteilung der sozialversicherungspflichtigen Einkommen oberhalb der Beitragsbemessungsgrenze unbekannt. Um jedoch die tatsächliche regionale Lohnverteilung abbilden zu können, ist ein Korrekturverfahren nötig. Zum Vergleich bietet sich hier die Lohnverteilung aus dem Mikrozensus oder der Einkommens- und Verbraucherstichprobe (EVS) an, die jedoch nur auf überregionaler Ebene (für Länder oder den Bund) und nicht auf Kreisebene erfolgen kann. Aber auch diese Einkommensverteilungen zeigen eine stetig abfallende Kurve für Einkommen oberhalb der Beitragsbemessungsgrenze.

Das beschriebene Verfahren kann sicherlich nur als eine Annäherung an die tatsächliche regionale Lohnverteilung der Bruttolohn- und Gehalts-

summe verstanden werden. Die regionalen Lohnverteilungen konvergieren bei den Einkommensklassen zwischen 100.000 € und 150.000 € gegen Null bedingt durch die Annahme einer stetigen Verteilungsfunktion über der Beitragsbemessungsgrenze. Somit werden auch mit der eingangs beschriebenen Lohnsummenkorrektur die Einkommen über der Beitragsbemessungsgrenze vermutlich noch leicht unterschätzt. Die Annahme einer stetigen Verteilung direkt oberhalb der Beitragsbemessungsgrenze ist nur insoweit für die folgenden 20 bis 40 Einkommensklassen gerechtfertigt, wenn jede Einkommensklasse eines Kreises eine bestimmte Mindestanzahl an Beschäftigten aufweist. Eine kreisspezifische zuverlässige Annäherung der oberen und unteren Einkommensklassen ist derzeit mit keiner analytischen Verteilungsfunktion möglich.

Will man jedoch Einkommensunterschiede von Regionen miteinander vergleichen (z.B. Nürnberg/Erlangen und Umland), so ist es unerlässlich, den institutionellen Fehler der Beitragsbemessungsgrenze zu bereinigen. Das beschriebene Verfahren stellt einen Ansatz hierfür vor.

Literatur

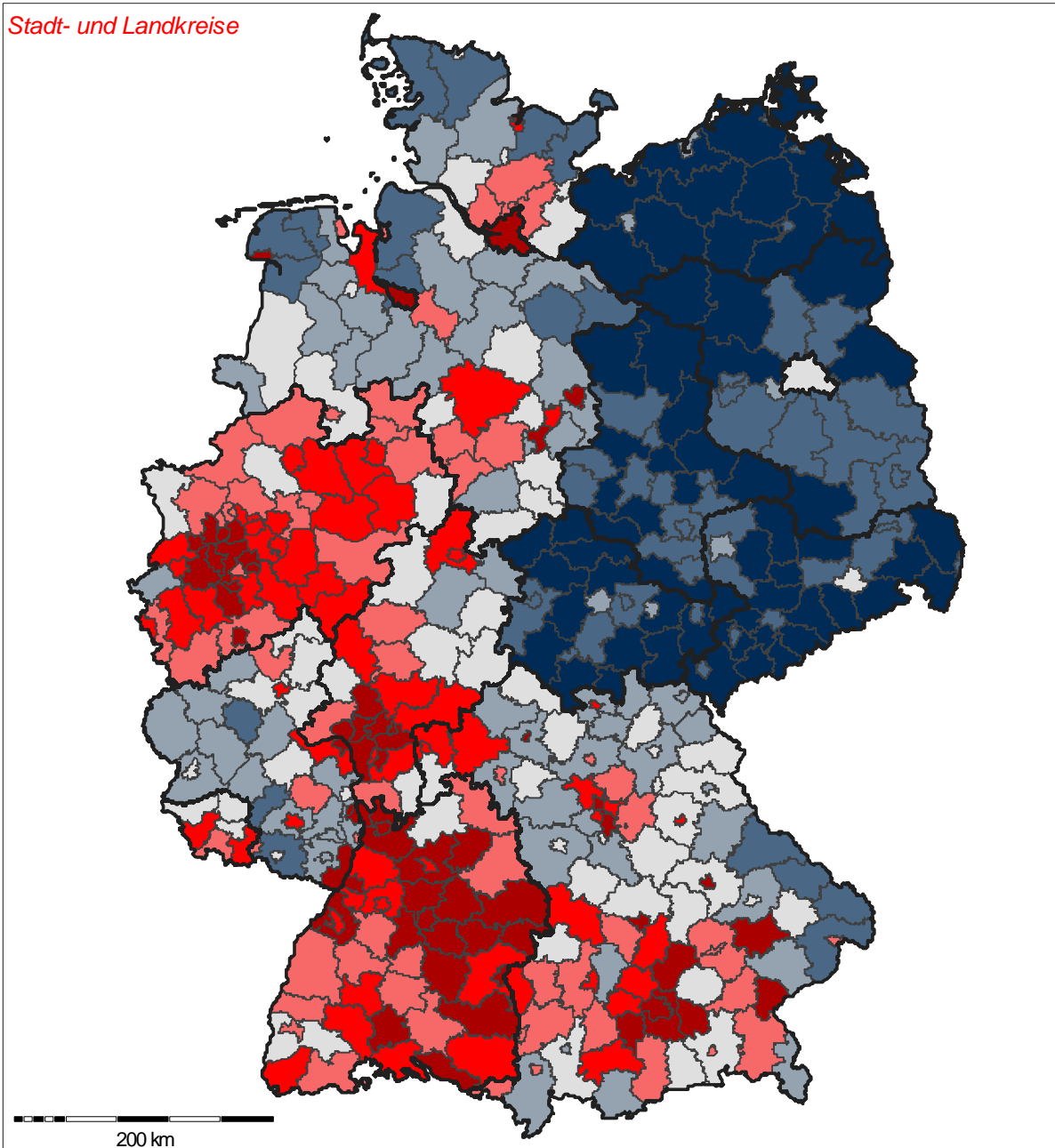
- Arntz, M./Feil, M./Spermann, A. (2003): Maxi-Arbeitsangebotseffekte oder zusätzliche Arbeitslose durch Mini- und Midi-Jobs? ZEW Discussion Paper No. 03-67.
- Blümle, G. (1975): Theorie der Einkommensverteilung, Springer.
- Deutscher Bundestag (2000): Neunundzwanzigster Rahmenplan der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ für den Förderzeitraum 2000-2003 (2004), Unterrichtung durch die Bundesregierung, Drucksache 14/3250.
- Fitzenberger, B./Reize, F. (2002): Quantilsregressionen der westdeutschen Verdienste: Ein Vergleich zwischen der Gehalts- und Lohnstrukturerhebung und der IAB-Beschäftigtenstichprobe. ZEW - Discussion Paper No. 02-79.
- Gartner, H./Rässler, S. (2005): Analyzing the changing gender wage gap based on multiply imputed right censored wages. IAB Discussion Paper No. 05/2005.
- Haag, G./Binder, J./Koller, M./Schwengler, B. (2002): Gutachtung zur Überprüfung der Fördergebiete, IAB-Gutachten Nr. 4 Teil 1 und Teil 2.
- Koller, M./Schwengler, B. (2000): Struktur und Entwicklung von Arbeitsmarkt und Einkommen in den Regionen. Beiträge zur Arbeitsmarkt- und Berufsforschung, BeitrAB 232. Bundesanstalt für Arbeit.

- Krupp, H.-J. (1968): Theorie der personellen Einkommensverteilung – Allgemeine Grundzüge und verteilungspolitische Simulation.
- Lankes, F. (1997): Entgeltverteilung der versicherungspflichtig Beschäftigten. Deutsche Rentenversicherung, 11-12, S. 703-718.
- Limpert, E./Stahl, W./Abbt, M. (2001): Lognormal distributions across the sciences: keys and clues. *BioScience*, 51(5), S. 341-352.
- Merz, J. (2002): Reichtum in Deutschland: Hohe Einkommen, ihre Struktur und Verteilung – Eine Mikroanalyse mit der Einkommensteuerstatistik für Selbständige und abhängig Beschäftigte. FFB-Diskussionspapier Nr. 36.
- Müller, H. (2005): Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik, Statistisches Bundesamt - Forschungsdatenzentrum-Arbeitspapier Nr. 1.
- Neal, D./Rosen, S. (2000): Theories of the Distribution of Earnings. In: *Handbook of Income Inequality*. Amsterdam, Kap. 7, S. 379–428.
- Roy, A.D. (1950): The distribution of earnings and of individual output. *The Economic Journal* 60 (1950), S. 489–505.
- Schwengler, B./Binder, J./Haag, G. (2006): Ausgewählte Regionalindikatoren zur Überprüfung von Vorranggebieten im Rahmen der regionalen Strukturpolitik. Gutachten im Rahmen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ im Auftrag der Senatsverwaltung für Wirtschaft, Arbeit und Frauen des Landes Berlin.
- Sheldon, G. (2002): Sind die Löhne einkommensmaximierender Personen vergleichbar? Das Roy-Modell im Rückblick. Arbeitspapier des Wirtschaftlichen Zentrums für Arbeitsmarkt- und Industrieökonomik, Basel, Schweiz.




Anhang

Abbildung 6: Korrigierter Bruttojahreslohn pro Kopf 2003


Stadt- und Landkreise



korrigierter Bruttojahreslohn pro Kopf [€]
(Einteilung in 7 gleiche Klassen)

	14.850 <= 18.200		23.800 <= 25.200
	18.200 <= 20.900		25.200 <= 27.000
	20.900 <= 22.500		27.000 <= 39.000
	22.500 <= 23.800		

West: 26.700€ Ost: 19.900€ BRD: 25.300€


 Steinbeis-Transferzentrum
 Angewandte Systemanalyse

IAB Institut für Arbeitsmarkt-
 und Berufsforschung

